VOLUME 61
NUMBER 3

WHOLE NO. 284
1947

Psychological Monographs

JOHN F. DASHIELL, *Editor*

A Comparison of the Use of Direct and Indirect Phrasing in Personality Questionnaires

By
ALBERT ELLIS

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Publications Office
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

I.

II.

III.

IV.

V.

Science Lib
Psychology
Closed
8-16-49
Shortage

TABLE OF CONTENTS

I. INTRODUCTION 1

II. EXPERIMENTAL MATERIALS, SUBJECTS AND DESIGN 4

III. THE DATA AND THEIR TREATMENT 7

IV. DISCUSSION 30

V. SUMMARY AND CONCLUSIONS 33

BIBLIOGRAPHY 35

ACKNOWLEDGEMENTS

I WISH gratefully to acknowledge the cooperation of all those who assisted me with this study: particularly my sponsor, Professor Percival M. Symonds, for his unwavering support, and Professors Helen M. Walker and Irving Lorge for their invaluable statistical and editorial guidance. Others to whom I am indebted for many useful suggestions include Professors Goodwin Watson, Laurance F. Shaffer, Robert Thorndike, Carl Warden, and Joseph Zubin.

I would also like to thank Gertrude Macarof and Rochelle M. Wexler for their valuable aid in facilitating the administration of the questionnaires employed in this study.

ALBERT ELLIS

New York, March 29, 1947

A COMPARISON OF THE USE OF DIRECT AND INDIRECT PHRASING IN PERSONALITY QUESTIONNAIRES

CHAPTER I

INTRODUCTION

WHEN Woodworth (28) constructed the Personal Data Sheet during the first World War, he started a fashion in personality questionnaires which has had an enormous influence on psychological and educational research and practice. In the course of the last thirty years, the popularity of paper and pencil personality tests has been impressively proved not only by their widespread use but also by the great number of studies devoted to their construction, standardization, reliability and validity.

One consequence of the extensive researches on conventional personality questionnaires has been the development, today, of instruments which are usually conceded to be highly reliable. In terms of validity, however, results have been far less decisive and promising. Material on the outcome of studies of the validity of personality inventories has been reviewed recently by the present author (4, 5), who summarized the validation data published between 1919 and 1946 on tests of the Woodworth, Bernreuter, and Thurstone type. The findings were as follows:

1. When used to distinguish between groups of problem and nonproblem children, personality questionnaires discriminated significantly in two experiments, discriminated with a questionable degree of significance in one, and were non-discriminative in six.

2. When utilized with delinquents, the questionnaires differentiated significantly between delinquent and non-delinquent groups in fifteen, to a questionable degree in six, and were non-differentiating in thirteen investigations.

3. When employed for purposes of psychiatric and psychological diagnosis, the inventories were adequate for group diagnosis in thirty-six, questionably adequate in nine, and inadequate in twenty-two studies.

4. When scores obtained by individuals taking personality questionnaires were compared to ratings given these individuals by their friends, associates, teachers, or supervisors, they were found to be closely related in twelve, related to a questionable degree in ten, and to a low degree in twenty-two investigations.

5. When intercorrelations between subjects' scores on two similar personality schedules purporting to measure the same traits were studied, they were found to be between .70 and 1.00 in nine, between .40 and .69 in eighteen, and between .00 and .39 in twenty-eight investigations.

6. When a review was made of experiments specifically designed to discover whether subjects taking personality questionnaires over-rated themselves or lied about their responses, it was discovered that thirty-six studies indicated that over-rating did take place, while only six showed that it did not.

On the basis of these findings, it was concluded that "group-administered paper and pencil personality questionnaires are of dubious value in distinguishing between groups of adjusted and maladjusted individuals, and . . . are of much less value in the diagnosis of individual adjustment or personality traits" (4, p. 426).

Personality inventories, however, have definite practical advantages. They are easy to administer and score; give results that may be objectively and statistically treated; may serve as a formalized interview; and at least give a picture of what the subjects *say* they feel and do. Because of these advantages, personality tests have maintained their popularity despite the recent growth of projective and non-paper-and-pencil techniques of assessing personality.

Is it possible to re-fashion personality questionnaires to eliminate some of their worst drawbacks and yet retain the advantages of ease of administration, scoring and interpretation? Several attempts in this direction have been made since Woodworth's original Personal Data Sheet was published. Thus, Laird (12) modified the YES-NO method of response by providing for several possible answers to each question. Bernreuter (3) introduced the inventory method of scoring a subject's answers to a single set of questions so that he receives ratings on several different traits. Maller (14) suggested stating the questions in a less personal form. Humm and Wadsworth (9), Washburne (27), and McKinley and Hathaway (7, 15) developed methods of estimating dishonesty and correcting for it. McKinley and Hathaway (7) employed a card-sorting, individual administrating, instead of the usual pencil-marking group administration. Adams and Lepley (1) favored innocuous questions that presumably would not embarrass respondents. Shipley and his associates (19) used a forced choice technique in the Personal Inventory.

All the foregoing modifications of the original Personal Data Sheet have added to the usefulness of personality questionnaires, but none has yet resulted in an instrument whose validity is unquestioned. Further experimentation in this connection would consequently seem desirable at present.

SPECIFIC PURPOSES OF THE PRESENT EXPERIMENT

The purpose of the present experiment was to study the effect of indirect phrasing in personality questionnaires. The question to be investigated was: Since the doubtful validity of conventional personality schedules may be partly attributable to respondents often discovering the intent of the questions, determining what the "good" or "correct" answers are, and then giving "good" responses, is it possible for inventory questions to be indirectly and less transparently stated so that schedules using them will be more effective for diagnostic and predictive purposes than are the present direct test forms?

Attempts to employ indirect questions in personality schedules are not entirely new. Sweet (23), Spencer (22), and Tuckman (26) devised questionnaires which included some indirectly phrased items. None of these investigators, however, specifically tried to compare the efficacy of direct and indirect personality questions utilized on the same subjects. The point of departure of the present experiment was that it studied the use of equated indirect and direct personality items in the same test battery; investigated the effect of these two kinds of personality questions on a behavior problem and a non-problem group of children; and tried to determine what was the effect of giving the direct and indirect questions in three different orders of presentation—a direct-indirect, alternate, and indirect-direct order. There were thus three major variables operating in the experiment: two *groups;* two kinds of *tests;* and three dif-

ferent *orders of presentation.*

The specific hypothesis of the experiment was: If conventional direct personality test forms, as well as indirect schedules derived from the same questions included on the direct forms, are given in a single battery to behavior problem and non-problem children: (1) the indirect test scores will be significantly different from the direct test scores; (2) groups of problem and non-problem children will react differently to the indirect and direct tests; and (3) presentation of the questions in three different orders—an indirect-direct, an alternate, and a direct-indirect order—will have a significant effect on the degree of test sophistication shown by the children in their responses.

Another way of stating the major premise of the experiment is to say that, under the given testing conditions, making use of the three major variables of tests, groups, and orders, it is hypothesized that significant differences in the subjects' mean test scores will result from the influence of every one of these variables.

Chapter II

Experimental Materials, Subjects and Design

MATERIALS

In order to obtain a schedule of direct questions that would be representative of the conventional tests widely used with children today, a frequency count was made of the items currently employed in the Woodworth-Mathews (28), Loofbourow-Keys (13), California (25), Pintner Aspects of Personality (17), Rogers (18), Hildreth (8), Maller (14), and Symonds (24) personality inventories. The thirty-six most popular questions were culled from these tests and were used as the basic direct questionnaire (Test D-1) employed in the experiment. (The complete questionnaire is printed in the Appendix.)

In order to see whether the manner of phrasing the direct questions themselves was of significant import, a second direct form (Test D-2) was constructed. In Test D-2 the same thirty-six questions used in Test D-1 were rephrased so that while D-1 items asked about a negative symptom, D-2 items asked about a positive one. Thus, where the third question on D-1 read, "I cry VERY OFTEN PRETTY OFTEN SELDOM NEVER," the equivalent item on D-2 read, "I am free from crying VERY OFTEN PRETTY OFTEN SELDOM NEVER." Whereas Test D-1 was scored so that the subjects received one point for each item answered VERY OFTEN or PRETTY OFTEN and no points for each one answered SELDOM or NEVER, Test D-2 was scored so that subjects received one point for each SELDOM or NEVER response and no points for each VERY OFTEN or PRETTY OFTEN answer.

The first of the two indirect test forms used in the experiment (Test I-1) was constructed by taking the same thirty-six questions utilized in Test D-1 and rephrasing them so that the subjects were asked to report how queer they considered other children who had symptoms like those described in the D-1 items. Thus, the I-1 version of the question on crying was, "Children who often cry are VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER." The theory behind this form of indirect questioning was that children possessing a symptom like frequent crying would tend to see other children who had this symptom as being A LITTLE QUEER or NOT AT ALL QUEER rather than as VERY QUEER or PRETTY QUEER. Preliminary experimentation showed, however, that the I-1 items were more meaningfully evaluated when a score of one point was given to each NOT AT ALL QUEER response and a score of no points to the VERY QUEER, PRETTY QUEER, and A LITTLE QUEER responses. This, consequently, was the scoring system employed for Test I-1.

The second indirect questionnaire employed in the experiment (Test I-2) was constructed by using, once again, the original thirty-six questions of Test D-1, but this time asking the subjects how many other children they thought possessed the symptoms described in the D-1 items. Thus, where on D-1 the subject was asked how often he cried, on the equivalent I-2 item he was asked to respond to the question, "VERY MANY PRETTY MANY FEW NO children often cry." The theory behind the I-2 rephrasings of the D-1 items was that the

subjects who actually possessed symptoms like crying frequently would tend to report that there were VERY MANY or PRETTY MANY children who also had these symptoms; while subjects who actually exhibited the symptoms seldom or never would tend to report FEW or NO other children as having them. Consequently, on I-2 items a score of one point was given for the VERY MANY or PRETTY MANY responses and a score of no points for the FEW or NO answers.

All four test forms, D-1, D-2, I-1 and I-2, were bound together in a single booklet and administered to the subjects at one sitting, in the manner explained below. The only additional material employed in the experiment was the Pintner Intermediate Test (16), a standard group test of intelligence which was routinely administered to the subjects several weeks before the administration of the personality tests.

SUBJECTS

The subjects of the experiment consisted of two groups of boys from regular seventh and eighth grade classes of three New York City junior high schools:

1. The behavior problem children—of which there were forty—were known as serious problem cases in their schools. This meant that each pupil had been brought to the attention of the school principal on several occasions and had been referred to the school psychologist for special guidance. While known as problem children, these forty subjects were not among the worst problems in the New York City schools, and were permitted to participate in regular school classes rather than special or segregated ones. There is little reason to assume that the problem group was a homogeneous one, since the offences against school discipline for which the members of the group were brought to the attention of the authorities were quite varied. Moreover, even when similar offences were committed, the reasons for them may have ranged from serious neurosis to a rather healthy aggressiveness.

2. The non-problem group consisted of 221 unselected school children in regular classes who were having no special behavior difficulties in school.

The test booklets were administered to the children by their own teachers during their home-room periods. All the children in each class were tested, the problem along with the non-problem subjects. Segregation of the two groups of papers was made only after they had been handed in and scored.

DESIGN OF THE EXPERIMENT

In designing the experiment, several considerations had to be taken into account:

1. The four personality test forms had to be given at equally spaced time intervals, and at intervals which would minimize the factor of intervening occurrences. This was accomplished by binding them together in a single test booklet which could be answered by all the subjects at one sitting. Since there were only 144 brief questions all told, the element of test fatigue was negligible.

2. The tests had to be presented in three different orders: (a) all the indirect questions followed by all the direct ones; (b) a page of indirect alternated with a page of direct questions; and (c) all the direct questions followed by all the indirect ones. This was done by dividing each test into four pages, each page containing nine printed questions. The pages were then assembled and bound in the three required orders, so that one-third of the children could be given in-

direct-direct, one-third alternate, and one-third direct-indirect booklets.

3. Perseveration tendencies on the part of the subjects had to be minimized. This was accomplished by assembling the sixteen pages of each booklet so that two pages of the same test form never followed each other. Thus, in the indirect-direct booklets, there would be a page of I-2 questions alternated with a page of I-1 questions for the first half of the battery; and a page of D-1 alternated with a page of D-2 questions for the last half of the booklet.

4. Randomization of the individual questions in each test had to be effected. This was done by numbering them on separate slips of paper, mixing them thoroughly in a box, and then picking them out individually and printing them in the order thus selected.

5. Randomization of the individuals and equalization of the numbers of subjects taking the three test orders had to be arranged. This was accomplished by stacking the booklets so that an indirect-direct battery was followed by an alernate one, which was in turn followed by a direct-indirect one, and so on. The booklets were then distributed to the subjects so that chance factors determined which test order any individual child received; while, in every class tested, approximately one-third of the children received a direct-indirect, one-third an alternative, and one-third an indirect-direct booklet. The fact that children who sat near each other had diferently arranged booklets also meant that attempts at copying were frustrated.

6. Environmental factors in the school situation had to be similar for both problem and non-problem subjects in order to effect maximal experimental control. This was done (as outlined above) by selecting problem children who attended regular school classes and who were therefore not responding under segregated or other conditions which might have effected the testing environment.

7. Randomization of individuals and equalization of numbers of subjects taking the three test orders had to be accomplished for problem children as well as non-problem children involved in the experiment. This was arranged by having the problem subjects answer the test battery in their regular classroom groups, exactly as did the non-problem children; so that no distinction was made between the two groups in the course of the test administration.

In brief: the experiment was so designed as to permit the four personality test forms to be given to the two experimental groups in three different orders of presentation without any systematic biases as to the arrangement of the questions, the subjects, the tests, the groups, or the orders.

Chapter III

The Data and Their Treatment

The basic data on the mean scores and variability obtained by 221 non-problem and 40 problem children on the four personality test forms used in the experiment are presented in Table 1.

From Table 1 it would appear that, for both the problem and non-problem groups: (1) Tests I-1 and I-2 differed from each other; (2) Tests D-1 and D-2 did not differ; (3) the two indirect tests differed from the direct ones; and (4) some order differences existed. The question is: Are these differences statistically significant?

Since three main variables were involved in the experiment—groups, tests, and orders—and since considerable interaction among these variables was possible, the analysis of variance technique, devised by Fisher (6) and ably expounded in this country by Snedecor (20, 21), seemed to be the most practical statistical method of treating the data of Table 1. Application of Bartlett's (2) test for the homogeneity of variance to the data revealed a sufficient degree of homogeneity to make the use of analysis of variance feasible, so it was accordingly employed.

Table 1

Mean scores and standard deviations. 221 non-problem and 40 problem children taking each of four forms of a personality test in three orders of presentation

Order of Presentation	Mean Score and Standard Deviation											
	Test I-1			Test I-2			Test D-1			Test D-2		
	N	M	S.D.	N	M	S.D.	N	M	S.D.	N	M	S.D.
No -Problem Children												
Indirect-direct	69	15.73*	7.44	69	13.33	6.55	69	4.29	3.92	69	3.75	3.53
Alternate	78	14.83	8.70	78	11.81	6.88	78	4.41	5.31	78	3.87	3.64
Direct-indirect	74	12.62	8.43	74	10.00	6.47	74	4.77	6.47	74	4.35	3.73
Problem Children												
Indirect-direct	13	16.53	6.07	13	12.46	5.11	13	6.54	6.84	13	6.92	4.96
Alternate	14	17.21	9.21	14	12.21	4.71	14	6.64	4.47	14	6.36	5.12
Direct-indirect	13	16.85	10.32	13	8.69	7.05	13	3.62	3.65	13	4.07	3.04

* All calculations were carried out to the third decimal place; but for ease of reading, only two decimal places are used in this and all succeeding tables.

The Problems Treated

Problem 1a. Were the subjects' mean scores on the four personality tests significantly affected by: Form of the test? Composition of the group taking the test? Order of presentation of the questions?

Table 2 presents the basic analysis of variance applied to the data of Table 1. Table 2 may be interpreted as follows: If the Mean Square of the individual scores within subclasses (37.09) is taken as the best estimate of the error variance

TABLE 2

Analysis of variance of mean scores, 221 non-problem and 40 problem children taking each of four forms of a personality test in three orders of presentation

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	$F_{.05}$	$F_{.01}$
Tests	3	21,095.61	7,031.87	189.61**	2.61	3.80
Groups	1	204.64	204.64	5.52*	3.85	6.66
Orders	2	397.38	198.59	5.36**	3.00	4.62
Interaction: Tests X Groups	3	171.63	57.21	1.54	2.61	3.80
Interaction: Tests X Orders	6	222.19	37.03	1.00	2.10	2.82
Interaction: Groups X Orders	2	49.40	24.70	.67	3.00	4.62
Interaction: Groups X Tests X Orders	6	428.52	71.42	1.93	2.10	2.82
Individual scores within subclasses	1020	37,827.49	37.09			
Total	1043	60,396.86				

* Significant.
** Highly significant.

among the individuals taking the four test forms, then there is considerably less than one chance in a hundred that a difference as large as the observed difference in the mean score among the four *tests* would be caused by chance factors; there is less than one chance in a hundred that a difference as large as the observed difference in the mean score among the three *orders* would be so caused; and there is only from one to five chances in a hundred that a difference as large as the observed difference in the mean score between the two *groups* would accidentally occur. Stated differently: if the individual score variation is taken as a valid estimation of error variance, it may be stated with great confidence that the obtained test and order differences, and with reasonable confidence that the observed group differences, did *not* arise by chance.

If, instead of using the variation among individual scores as the best estimate of variance, the standard error of estimate is calculated by combining the appropriate mean squares of first-order interactions, second-order interactions, and individual scores, and dividing by their combined degrees of freedom, the F-values for tests, groups and orders respectively become 188.37, 5.56, and 5.32—which are almost identical with the values in Table 2. This analysis of variance, evidently, is so stable that the F-values maintain their significance even when their base is changed—a fact which should enhance confidence in the significance of the observed mean differences.

Another way of interpreting the three significant F-values of Table 2 is to say that, with the influence of groups and orders eliminated, the mean scores of the four tests were significantly different. With the effect of tests and orders eliminated, the group differences were significant. Finally, with effects of tests and groups eliminated, order of presentation differences were significant.

Problem 1b. Did presentation of all four test forms in a single battery pro-

duce significant interaction effects?

Since the first- and second-order interaction *F*-values of Table 2 are all low, and none of them reaches significance at even the 5% level of confidence, there is no evidence that significant degrees of interaction among tests, groups, or orders took place.

Problem 1c. Were there significant differences between scores obtained on: The first and second direct tests? The first and second indirect tests? The direct and indirect tests?

Table 2 must be analyzed into smaller segments to provide more detailed conclusions to several problems of the experiment. Thus, although the *F*-value of 189.61 for tests shows that the four tests differed significantly among themselves, it does not clearly indicate whether D-1 mean scores differed from those of D-2, I-1 from I-2, nor D-1 and D-2 from I-1 and I-2. However, if the mean square for tests is appropriately separated into the three degrees of freedom of which it is comprised, these questions may be answered:

1. Between the mean scores of Tests D-1 and D-2, *F*-values of 2.45 and .02 are obtained, respectively, for non-problem and problem subjects. Neither of these *F*-ratios is significant at the .05 level of confidence.

2. Between Tests I-1 and I-2 highly significant *F*-ratios of 49.07 and 65.06 are obtained, respectively, for non-problem and problem children. Apparently, the two direct tests did not differ significantly from each other, but the two indirect ones did.

3. Between the two direct tests, on the one hand, and the two indirect schedules, on the other hand, *F*-ratios are found to be, for non-problem and problem children respectively, 101.48 and 8.05, both of which are highly significant. Thus, the two direct tests differed distinctly from the two indirect forms.

Problem 1d. Which of the four tests significantly differentiated the problem from the non-problem groups of subjects?

While Table 2 indicates that there were significant differences between the mean scores of the problem and non-problem children, it does not specifically show which of the four tests accounted for these differences. Accordingly, the data were arranged so that analyses of variance could be done by individual tests. These anlayses are presented in Table 3

From Table 3 it may be seen that the *F*-ratios for group differences for Tests I-1 (3.08), I-2 (.02) and D-1 (2.09) are not significant. However, that of Test D-2 (7.26) is quite significant. In other words, on Tests I-1, I-2, and D-1 there is a good chance that differences as large as the observed differences between the mean scores of problem and non-problem children could have arisen purely by accident. But there is less than one chance in a hundred that the obtained group difference on Test D-2 was an accidental one.

Problem 1e. Did order of presentation of the questions make a significant difference on test scores?

Table 2 indicates that order of presentation of the questions effected significant differences in mean test scores. Table 3 indicates that when problem and non-problem groups were considered together, significant order differences occurred with Test I-2, but not with Tests I-1, D-1, and D-2. When analyses of variance are performed on the mean test scores with groups and tests considered separately, the *F*-ratios shown in

TABLE 3

Analysis of variance of mean scores. 221 non-problem and 40 problem children taking each of four forms of a personality test in three orders of presentation

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	$F_{.05}$	$F_{.01}$
Test I-1						
Groups	1	213.13	213.13	3.08	3.89	6.76
Orders	2	312.28	156.14	2.26	3.04	4.71
Interaction:						
Groups X Orders	2	59.65	29.73	.43	3.04	4.71
Individual scores	255	17,621.94	69.11			
Total	260	18,207.00				
Test I-2						
Groups	1	9.47	9.47	.02	3.89	6.76
Orders	2	495.77	247.83	5.51**	3.04	4.71
Interaction:						
Groups X Orders	2	19.72	9.86	.22	3.04	4.71
Individual scores	255	11,217.81	43.99			
Total	260	11,742.77				
Test D-1						
Groups	1	43.39	43.39	2.09	3.89	6.76
Orders	2	1.12	.56	.03	3.04	4.71
Interaction:						
Groups X Orders	2	85.81	42.91	2.16	3.04	4.71
Individual scores	255	5,171.69	20.28			
Total	260	5,302.01				
Test D-2						
Groups	1	110.29	110.29	7.26**	3.89	6.76
Orders	2	.82	.41	.02	3.04	4.71
Interaction:						
Groups X Orders	2	73.32	36.66	2.42	3.04	4.71
Individual scores	255	3,865.26	15.16			
Total	260	4,029.69				

** Highly significant.

Table 4 are obtained. In these analyses, the variance among the mean scores in the three orders (indirect-direct, alternate, and direct-indirect) are compared with the estimate of error variance obtained by considering the variation among the scores of the individuals taking each test order.

If Table 4 is read in conjunction with Table 1 the following points may be noted:

1. Non-problem children taking Test I-1 obtained a higher (that is, worse) mean score on the indirect-direct (15.73) than on the alternate (14.83) or the direct-indirect (12.62) orders of presentation. This increase in indirect-direct mean score is statistically significant.

2. Non-problem children taking Test I-2 also obtained higher scores on the indirect-direct (13.33) than on the alternate (11.81) or the direct-indirect (10.00) orders. This, too, is a statistically significant increase.

3. Non-problem children taking test D-1 obtained a lower (or better) mean score on the indirect-direct (4.20) than the alternate (4.41) or the direct-indirect

TABLE 4

Order of presentation *F*-Ratios (order of presentation variance/error variance). 221 non-problem and 40 problem children taking each of four forms of a presonality test in three orders of presentation

Test	Observed Value of *F*	$F_{.05}$	$F_{.01}$
Non-Problem Children			
Test I-1	3.09*	3.04	4.71
Test I-2	5.60**	3.04	4.71
Test D-1	2.02	3.04	4.71
Test D-2	2.34	3.04	4.71
Problem Children			
Test I-1	.02	3.25	5.21
Test I-2	.98	3.25	5.21
Test D-1	1.45	3.25	5.21
Test D-2	1.47	3.25	5.21

* Significant.
** Highly significant.

(4.77) order. This decrease is not statistically significant.

4. Non-problem children taking test D-2 again obtained a lower mean score on the indirect-direct (3.75) than on the alternate (3.87) or the direct-indirect (4.35) order of presentation. This, too, is not a significant difference.

5. Problem children taking the I-1 test obtained a lower mean score on the indirect-direct (16.54) than on the alternate (17.21) or the direct-indirect (16.85) order. But problem children taking Tests I-2, D-1, and D-2 obtained higher mean scores on the indirect-direct than they generally did on the alternate and direct-indirect orders of presentation. However, none of these differences are significant.

The data of Tables 1 and 4 seem to confirm the hypothesis of test sophistication. For when the direct questions were printed prior to the indirect ones, non-problem subjects made significantly lower, or better, scores on both indirect tests. But when the indirect questions were printed first, subjects made somewhat (though not statistically reliable) lower, or better, scores on both direct tests. It would seem that presenting the direct questions prior to the indirect ones enabled non-problem children to discover the intent of the latter and answer them in a favorable manner. Presenting indirect questions prior to direct ones also produced a degree of test sophistication on the latter, but not enough to make the mean score differences statistically significant. The fact that, in all four tests taken by non-problem children, the mean scores on the alternate order of presentation were always in between those obtained on the indirect-direct and direct-indirect orders lends additional weight to the theory that a measurable degree of test sophistication took place.

Concerning the behavior problem children, the evidence on order of presentation differences is conflicting and baffling. It is probable that since so few problem subjects were employed in each sub-grouping, and since the sub-groups may have differed widely among themselves, the observed order differences here—none of which reaches statistical significance—are meaningless.

Problem 2a. How did the internal consistence or reliability of the four tests affect the subjects' scores?

It is important to know the internal consistency or reliability of each of the four tests to see if lack of test reliability may have accounted for some of the observed differences. As Jackson and Ferguson (10) have pointed out, what does or does not constitute a true test reliability is still a matter of controversy. Test-retest, split-half, internal consistency and other methods commonly used

for obtaining reliabilities each has its advantages and disadvantages. In the present study, the Kuder-Richardson (11) formula for measuring internal consistency was employed—because the four test forms were administered only once and because this formula gives a single unique estimate of internal consistency or reliability.

Table 5 presents the internal consistency quotients obtained by applying the Kuder-Richardson formula No. 20 to the original score data of the experiment. It can be seen that while none of the correlations in Table 5 is exceptionally high, those of the non-problem children are all significantly different from zero at the 1% level of confidence, according to Fisher's (6) test for the significance of correlations. Those of the problem children, however, are not as high as those of the non-problem subjects.

The internal consistency quotients in Table 5 may be converted into z-values (Fisher's transformation) and tested for significant differences among them. When this was done, and an analysis of variance performed upon the z-values thus obtained, using the second-order interaction among tests, groups, and assemblies as the error of estimate, non-significant *F*-ratios were found for tests (.92), groups (.18) and orders (.22).* It would seem, therefore, that there were no significant test, group or order differences in the internal consistency quotients. Apparently, the observed mean score differences were not attributable to differences in the internal consistency of the four test forms.

Problem 2b. Were the obtained differences in mean scores attributable to lack of item stability on the four tests?

To discover if the tests possessed a sufficient degree of item stability to warrant the assumption that the obtained test, group, and order mean score differences were not merely caused by the use of unstable testing instruments, an item analysis was done of each of the four tests used with non-problem chil-

TABLE 5

Internal consistency or reliability coefficients (Kuder-Richardson). 221 non-problem and 40 problem children taking each of four forms of a personality test in three orders of presentation

Test	Internal Consistency Quotient of Non-Problem Children			Internal Consistency Quotient of Problem Children		
	Order of Presentation			Order of Presentation		
	Indirect-Direct (N = 69)	Alternate (N = 78)	Direct-Indirect (N = 74)	Indirect-Direct (N = 13)	Alternate (N = 14)	Direct-Indirect (N = 13)
I-1	.90	.92	.86	.76	.91	.93
I-2	.89	.90	.76	.70	.62***	.88
D-1	.82	.91	.75	.94	.83	.75
D-2	.78	.79	.78	.84	.94	.50***

*** All the correlations in the table are statistically significant from zero at the 1% level of confidence except the two starred ones.

To discover if the tests possessed a sufficient degree of item stability to warrant the assumption that the obtained test, group, and order mean score differences were not merely caused by the use of unstable testing instruments, an item analysis was done of each of the four tests used with non-problem chil-

* This analysis of variance of the z-values, as well as those made in the next few pages, is based on the assumption of equal reliability for the original correlations and therefore is not to be viewed as too precise a test of significance of test, group, and order differences. While it is the best that can be made under the circumstances, its results should be taken as approximations rather than as exact determinations.

dren. A similar analysis was not done for the problem children's responses because the small number of problem children employed in the experiment made it impractical.

In the item stability analysis, the biserial correlation was found between total test score and each item response. The obtained correlations, classified according to order of presentation of the tests, are shown in Table 6. An examination of Table 6 reveals the following:

TABLE 6

Bi-serial correlations between responses to individual test items and total test scores. 221 non-problem children taking each of four forms of a personality test in three different orders of presentation

Item	Bi-serial Correlation Coefficient Between Item and Total Score											
	Test I-1			Test I-2			Test D-1			Test D-2		
	I-D[a]	ALT[b]	D-I[c]	I-D	ALT	D-I	I-D	ALT	D-I	I-D	ALT	D-I
1.	.30	.87	.57	.40	.90	.70	.88	.36	.53	.79	.77	.63
2.	.48	.85	.79	.32	.42	.52	.38	.90	.90	.14	.28	.89
3.	.75	.61	.89	.50	.38	.58	.42	.48	.38	.50	.74	.41
4.	.59	.89	.88	.52	.65	.60	.57	.41	.74	.41	.74	.69
5.	.72	.74	.81	.53	.78	.07	.37	.55	.81	.92	.88	.95
6.	.69	.88	.92	.77	.35	.81	.70	.85	.80	.48	.28	.41
7.	.38	.80	.86	.68	.65	.67	.78	.75	.44	.50	.92	.69
8.	.53	.49	.82	.39	.71	.73	.70	.94	.06	.31	.89	.41
9.	.49	.57	.88	.45	.68	.68	.15	.75	.75	.22	.75	.95
10.	.50	.55	.87	.66	.78	.63	.38	.88	.92	.46	.42	.89
11.	.28	.63	.63	.50	.50	.65	—[d]	.83	.86	—	.91	.88
12.	.43	.91	.91	.59	.49	.57	.18	.89	.91	.45	.41	.32
13.	.45	.64	.78	.82	.39	.43	.73	.60	.27	.53	.88	.46
14.	.69	.64	.71	.44	.48	.67	.84	.61	.49	.51	.70	.33
15.	.74	.82	.68	.32	.81	.57	.91	.93	.11	.60	.91	.88
16.	.62	.55	.89	.56	.71	.88	.57	.77	.65	.29	.93	.63
17.	.73	.72	.92	.42	.48	.82	.85	.89	.58	.37	.08	.59
18.	.29	.40	.63	.43	.08	.58	.38	.53	.27	.49	.70	.63
19.	.73	.71	.96	.76	.83	.69	.06	.45	.15	.41	.31	.50
20.	.54	.30	.47	.78	.16	.77	.59	.75	.62	.59	.82	.48
21.	.48	.86	.89	.92	.53	.74	.87	.90	.93	.55	.91	.72
22.	.52	.89	.79	.55	.43	.71	.86	.48	.33	.58	.60	.27
23.	.55	.74	.88	.45	.55	.55	.69	.92	.58	.43	.28	.84
24.	.61	.86	.83	.64	.89	.60	.62	.78	.74	.50	.55	.62
25.	.50	.63	.66	.61	.87	.79	.31	.94	.09	.48	.67	.79
26.	.35	.73	.72	.60	.44	.52	.34	.52	.92	.30	.11	.89
27.	.70	.76	.70	.92	.92	.77	.50	.95	.42	.33	.77	.19
28.	.59	.87	.72	.92	.43	.70	.51	.70	.74	.17	.70	.86
29.	.59	.71	.79	.71	.89	.61	.17	.39	.53	.20	.68	.30
30.	.39	.55	.24	.33	.40	.54	.39	.66	.75	.65	.40	.18
31.	.67	.46	.76	.74	.82	.46	.59	.90	.58	.27	.88	.89
32.	.58	.50	.57	.21	.40	.48	.37	.42	.40	.39	.79	.48
33.	.59	.76	.53	.53	.88	.68	.92	.93	.71	.80	.92	.57
34.	.60	.77	.62	.85	.43	.64	.37	.57	.85	.63	.52	.73
35.	.57	.35	.68	.71	.44	.63	.23	.94	.22	.37	.95	.17
36.	.58	.46	.55	.39	.62	.60	.20	.61	.51	.14	.40	.47
Average	.58	.71	.78	.63	.64	.63	.54	.79	.67	.46	.73	.68

[a] Indirect-direct order of presentation.
[b] Alternate order of presentation.
[c] Direct-indirect order of presentation.
[d] Blank spaces indicate that there were no responses to one of the categories, all the individuals answering the item in one way.

1. All the correlations are positive—which indicates that every item in the four tests contributed something to the schedule in which it was used. Most of the correlations are above .42, and are consequently significantly different from zero at the 1% level of confidence.

2. When the correlations in Table 6 are transformed into z-values and average correlations calculated for each order of presentation, the averages are all significantly different from zero at the 1% level of confidence. Evidently, then, the tests employed in the experiment were not composed of haphazardly thrown together items, but did possess some degree of item stability.

The biserial correlations in Table 6 may be transformed back into z-values and analyzed for significant differences among them. When this was done, using the mean square of the individual items as the estimate of error variance, analyses of variance produced *F*-ratios of 8.00 for tests (3 and 420 degrees of freedom) and 62.00 for orders (2 and 420 degrees of freedom), both of which are highly significant. That is to say: there were significant differences among both tests and orders, thus confirming the test and order mean score differences previously discovered.

Problem 3. What was the effect of the intelligence quotients of the subjects on their personality test scores?

To uncover the possible effect of intelligence quotients on personality test scores, the Pintner Intermediate Test of Mental Ability was given to the subjects. On this test, the non-problem children had an average intelligence quotient of 109.74 and the problem children of 100.42. However, the problem children's mean chronological age was 13.82 against the non-problem children's mean of 12.84. Mean mental ages of the problem children were found to be, for the indirect-direct, alternate, and direct-indirect orders respectively, 13.85, 13.97 and 13.86; while those of the non-problem subjects for the same respective orders, were 14.11, 14.21, and 13.90. When tested by Chi-square analysis, these were not found to be significantly different, since Chi-square equalled only .84 for two degrees of freedom.

The correlations between the intelligence quotients of the individual children and their scores on the four tests are shown in Table 7. Table 7 indicates that whereas in many instances the non-problem children showed a significant negative degree of relationship between their intelligence quotients and test scores (that is, the higher their intelligence, the lower and more favorable their scores), there was only one notable instance of this occurring with the problem children.

To test the significance of the group, test, and order differences in the correlations listed in Table 7, these were converted into z-values and an analysis of variance performed. Using the mean square of the second-order interaction (tests X groups X orders) as the estimate of error variance, these *F*-ratios were found: for groups, 4.11 (1 and 6 degrees of freedom); for tests, 1.55 (3 and 6 degrees of freedom); and for orders, 2.84 (2 and 6 degrees of freedom). All these *F*-ratios are non-significant, indicating that the observed mean score differences among tests, groups, and orders probably were not caused by differential intelligence quotient effects.

Problem 4. What were the intercorrelations between: The direct and indirect tests? The two direct tests? The two indirect tests?

The intercorrelations between the scores of non-problem and problem chil-

TABLE 7

Correlation coefficients between intelligence quotients and scores, 221 non-problem and 40 problem children taking each of four forms of a personality test in three orders of presentation

Relationship	Correlation Coefficient of Non-Problem Children			Correlation Coefficient of Problem Children		
	Order of Presentation			Order of Presentation		
	Indirect-Direct (N = 69)	Alternate (N = 78)	Direct-Indirect (N = 74)	Indirect-Direct (N = 13)	Alternate (N = 14)	Direct-Indirect (N = 13)
Intelligence Quotients and I-1 Test Scores	.04	−.41**	−.43**	.25	.44	.09
Intelligence Quotients and I-2 Test Scores	−.26*	−.49**	−.64**	.26	−.22	−.66*
Intelligence Quotients and D-1 Test Scores	−.44**	.01	−.31**	−.02	.14	.16
Intelligence Quotients and D-2 Test Scores	.06	−.14	−.47**	.10	.06	−.45

* Significantly different from zero at the 5% level of confidence.
** Significantly different from zero at the 1% level of confidence.

dren on all four test forms are presented in Table 8. From an examination of Table 8 these points may be noted:

1. The two indirect tests, except for the indirect-direct order taken by the non-problem children, showed only a low degree of intercorrelation.

2. The degree of correlation between the indirect and direct tests was not very high for either the problem or non-problem subjects.

3. The degree of relationship between Tests D-1 and D-2 was reasonably high for the non-problem, but not for the problem children.

4. The indirect-direct order of presentation generally resulted in higher test intercorrelations than did the direct-indirect order.

When the correlations in Table 8 were converted to z-values and an analysis of variance done upon them, using the mean square of the second-order interaction as estimated error variance, these

TABLE 8

Intercorrelation coefficients for four personality tests taken by 221 non-problem and 40 problem children in three orders of presentation

Relationship	Intercorrelation Coefficient of Non-Problem Children			Intercorrelation Coefficients of Problem Children		
	Order of Presentation			Order of Presentation		
	Indirect-Direct (N = 69)	Alternate (N = 78)	Direct-Indirect (N = 74)	Indirect-Direct (N = 13)	Alternate (N = 14)	Direct-Indirect (N = 13)
Tests I-1 and I-2	.75**	.12	−.09	.47	−.23	.17
Tests I-1 and D-1	−.01	.17	.33**	.54	.25	−.23
Tests I-1 and D-2	.09	.19	.26*	.66*	.16	−.40
Tests I-2 and D-1	.24*	.48**	.22	.66*	.63*	.35
Tests I-2 and D-2	.37**	.52**	.04	.61*	.71**	.31
Tests D-1 and D-2	.83**	.77**	.67**	.97**	.57*	.38

* Significantly different from zero at the 5% level of confidence.
** Significantly different from zero at the 1% level of confidence.

F-ratios were obtained: relationships, 4.13 (5 and 10 d.f.); groups, .56 (1 and 10 d.f.); and orders, 6.95 (2 and 10 d.f.). These *F*-ratios indicate significant order differences at the 5% level of confidence; but non-significant group differences.

Problem 5. How significantly did the items on the four tests distinguish between non-problem and problem subjects?

To discover which test items distinguished significantly between problem and non-problem groups, an item analysis was performed, using the Chi-square method as modified by Yates (29) for use with small cell figures. Specifically, for the 36 items on each of the tests, the number of non-problem and problem subjects answering the item positively and negatively were recorded in two-by-two tables. A Chi-square test for homogeneity of each two-by-two table was then made. If a sufficiently large Chi-square figure (representing a probability of .05 or less) was obtained, it was assumed that homogeneity did not exist between the responses of the two groups of subjects, and the item was considered to distinguish significantly between them.

Table 9 lists Chi-square values only for those items which differentiated significantly between non-problem and problem children on the four tests employed in the experiment, classified according to order of presentation. For the purposes of readability, all insignificant Chi-square values are omitted from the table.

Observation of Table 9 reveals the following:

1. Not one of the four tests contained many items which significantly differentiated non-problem from problem children. Out of 108 possible items for each test (36 test items given in three different orders), Test I-1 included only 12 significant ones; Test I-2, 4; Test D-1, 10; and Test D-2, 17. Statistically considered, this is a meagre showing since about five items of each test, on a chance basis alone, should be expected to show Chi-square values as large as those tabulated for the .05 level of confidence.

2. The indirect tests included fewer differentiating items than did the direct tests. However, when the number of direct and indirect test discriminations were treated by Chi-square (that is, 16 out of a possible 216 compared to 27 out of 216), Chi-square was found to equal .002 for one degree of freedom. The difference is not significant.

3. Only in five out of fifteen instances were significant direct test items also significant on one or more of the orders of the indirect tests. On the other hand, seven of the ten significant D-1 items were also significant on one or more orders of the D-2 test. This substantiates previous findings that the two direct tests differed from the indirect ones but were similar to each other.

4. Some order of presentation differences are evident in Table 9. Test I-1 showed only one significant item value for its indirect-direct administration; but as many as eight on its direct-indirect. Test I-2 had too few significant items in any presentation order for realistic comparison. Test D-1 displayed four significant items on the indirect-direct order but only one on the direct-indirect. Test D-2 also showed a drop from ten significant items on the indirect-direct order to two on the direct-indirect.

5. Analysis of the content of the items which were discriminating in Table 9 shows no prominent patterning. With the direct tests, however, there is an interesting finding. The subject mater of the eight questions which discriminated on one or more of the presentation

TABLE 9

Chi-square values indicating the ability of individual test items to discriminate significantly between non-problem and problem groups of children taking each of four forms of a personality test in three orders of presentation

Item	Obtained Chi-square Value											
	Test I-1			Test I-2			Test D-1			Test D-2		
	I-D[a]	Alt.[b]	D-I[c]	I-D	Alt.	D-I	I-D	Alt.	D-I	I-D	Alt.	D-I
1.	—[d]	—	16.9	—	—	—	—	—	—	—	—	—
2.	—	—	3.7	—	—	—	—	—	—	—	—	—
3.	—	—	—	—	—	—	—	—	—	—	—	—
4.	4.4	—	17.5	—	—	—	—	—	—	3.7	7.4	—
5.	—	—	—	—	—	—	—	20.8	—	—	9.3	5.1
6.	—	—	—	—	—	—	—	—	—	—	—	—
7.	—	—	—	—	—	—	—	—	—	—	—	—
8.	—	—	—	—	—	—	—	—	—	—	—	—
9.	—	—	6.9	4.4	—	—	10.6	—	—	4.6	—	—
10.	—	—	—	—	—	—	—	21.2	—	5.2	5.5	—
11.	—	—	—	—	—	—	—	—	—	5.4	—	—
12.	—	—	—	—	—	—	—	3.7	—	9.3	—	—
13.	—	—	—	—	—	—	5.9	—	—	7.3	—	—
14.	—	—	—	—	—	—	—	—	—	—	—	—
15.	—	3.7	4.2	—	—	—	—	—	—	—	—	3.7
16.	—	—	—	—	—	—	—	—	—	—	—	—
17.	—	—	—	—	—	—	4.7	—	—	—	—	—
18.	—	—	—	—	—	—	—	—	6.0	—	4.0	—
19.	—	—	3.8	—	—	—	—	—	—	—	—	—
20.	—	—	—	—	—	—	—	—	—	—	—	—
21.	—	—	—	—	—	—	—	—	—	—	—	—
22.	—	—	—	—	—	—	—	—	—	—	—	—
23.	—	—	—	—	—	—	—	—	—	—	—	—
24.	—	—	—	—	—	—	—	—	—	—	—	—
25.	—	—	—	4.2	—	—	—	—	—	—	—	—
26.	—	4.0	—	—	—	—	—	14.7	—	7.1	16.2	—
27.	—	—	—	—	—	—	8.9	—	—	4.2	—	—
28.	—	—	3.7	—	—	—	—	—	—	—	—	—
29.	—	—	—	—	—	—	—	—	—	—	—	—
30.	—	—	—	—	—	—	—	—	—	—	—	—
31.	—	4.4	—	—	—	—	—	5.5	—	—	—	—
32.	—	—	—	—	—	—	—	—	—	—	—	—
33.	—	—	—	—	—	—	—	—	—	8.9	—	—
34.	—	—	3.8	8.5	—	—	—	—	—	—	—	—
35.	—	—	—	—	5.5	—	—	—	—	—	—	—
36.	—	—	—	—	—	—	—	—	—	7.1	—	—

[a] Indirect-direct order of presentation.
[b] Alternate order of presentation.
[c] Direct-indirect order of presentation.
[d] Blank spaces represent non-significant Chi-square values; all printed Chi-square values are significant at least at the 5% level of confidence.

orders of Tests D-1 and D-2 was as follows:

No. 5 Finds it hard to do his school work right.

No. 9 Feels other members of family are treated better than he.

No. 10 Finds school a hard place to get along in.

No. 12 Finds it hard to speak out in class recitations.

No. 13 Feels that his teachers treat him badly or unfairly.

No. 18 Feels tired when he wakes up

in the morning.
No. 26 Dislikes school.
No. 27 People find fault with him.

Five of these eight questions are concerned with school activities; and these happen to be the *only* questions on school activities in the entire test. Thus, the problem children gave answers which discriminated them significantly from the non-problem children when specifically questioned about their school activities.

Problem 6. Did item identity take place between the same questions used in: Tests D-1 and D-2? Tests D-1 and I-1? Tests D-1 and I-2?

Since Test D-1 was the basic instrument used, and the other three forms were derived by rephrasing its questions, it would be desirable to know whether item identity occurred between D-1 questions and corresponding D-2, I-1 and I-2 items. To answer this problem, two-by-two tables were set up showing how the same subjects responded to corresponding items in Test D-1 and in each of the other tests. Thus, if the 74 non-problem children taking the direct-indirect order of the four tests gave 50 positive responses to a given D-1 item and 24 negative responses; if these same subjects gave 44 positive and 30 negative responses to the corresponding I-1 item; if 20 of the 30 children answering the I-1 item negatively also answered the D-1 item negatively; and if 40 of the 44 children answering the D-1 question positively also answered the corresponding I-1 question positively; then the two-by-two table set up for determining possible item identity from this D-1 to I-2 item would look like this:

TEST I-1

	40 Positive on D-1 and Positive on I-1	10 Positive on D-1 and Negative on I-1	50 Total Positive on D-1
Test D-1	4 Negative on D-1 and Positive on I-1	20 Negative on D-1 and Negative on I-1	24 Total Negative on D-1
	44 Total Positive on I-1	30 Total Negative on I-1	74 Total Positive and Negative

The specific point at issue in this two-by-two table is whether, out of the 24 children who answered the D-1 item negatively, the proportion who also answered the I-1 item negatively (20/24) is greater than the proportion, out of the 50 children who answered the D-1 item positively, who also answered the I-1 item negatively (10/50). For if it can be shown that a significantly higher proportion of the negative than the positive scorers on a D-1 item also gave negative answers to the corresponding I-1 item, then presumably a distinct identity existed between the D-1 and I-1 responses to this item. In other words, if the hypothesis of item identity is correct, subjects who admit the possession of symptoms like frequent crying on Test D-1 will have a pronounced tendency to report that children who cry frequently are not at all queer (Test I-1); that there are many such children (Test I-2); and

that they are seldom or never free from crying (Test D-2).

Statistically, the two-by-two tables used for item identity analysis were tested by the Chi-square formula which includes Yates' correction:

$$\chi^2 = \frac{[(ad) - (bc) - N/2]^2 N}{(a+b)\,(c+d)\,(a+c)\,(b+d)}$$

When this formula is employed with the figures in the two-by-two table presented above, Chi-square is found to equal 38.4 for one degree of freedom; and the null-hypothesis—that this is a homogeneous distribution and that item identity does not take place—may be confidently rejected.

In accordance with this procedure, the 36 questions on each of the four tests were subjected to an analysis to determine whether item identity occurred from Tests D-1 to D-2, D-1 to I-1, and D-1 to I-2. The results of this analysis for the non-problem children's responses are presented in Table 10. Only significant Chi-square values are listed in the table. The *total* columns were obtained by adding independent Chi-squares for the three order columns under each test heading and using three instead of one degree of freedom.*

Several observations may be made from Table 10:

1. There was considerable item identity between non-problem children's responses to Test D-1 and D-2 items. But D-1 to I-1 and D-1 to I-2 identities were far less than the D-1 to D-2 identities.

2. There was decidedly more item identity from D-1 to I-2 than from D-1 to I-1. Children who admitted the possession of unfavorable personality symptoms were more willing to see many other children as possessing these symptoms than to call others who possessed them "queer."

3. More questions showed item identity than were significantly discriminating (according to Table 9) between problem and non-problem children.

4. The subject matter of the items which showed the most consistent identity from D-1 to D-2, D-1 to I-1, and D-1 to I-2 was as follows:

No. 4 Feels like running away from home.
No. 13 Feels that his teachers treat him badly or unfairly.
No. 22 Is afraid of high places.
No. 26 Dislikes school.
No. 32 Finds it hard to become a leader in games with other children.
No. 36 Is afraid of his father.

These items covered the whole range of questions used in the tests. There does not seem to be any consistent patternings among them.

When an item analysis was done with the test results of the problem subjects, Chi-square values shown in Table 11 were obtained. This time, however, the three orders of presentation for each test were combined, since—as the analysis of variance of mean scores showed—there were no significant order differences for the problem children. Moreover, there were too few problem subjects taking each order to allow for effective subdivisional calculations.

Examination of Table 11 reveals the following:

1. Item identity from Test D-1 to D-2 was not as prominent as it was with the non-problem subjects; but D-1 to I-2

* Strictly speaking, this is not entirely accurate, since the use of Yates' correction, according to Snedecor (21, p. 193) interferes with the additive nature of Chi-square. The figures in the *total* columns of Table 10 may therefore be taken only as suggestive.

TABLE 10

Chi-square values indicating a significant degree of item identity from the items on one personality test to the equivalent items on three other personality tests taken by 221 non-problem children in three orders of presentation

Item	Obtained Chi-square Value											
	Test D-1 to D-2				Test D-1 to I-1				Test D-1 to I-2			
	I-D[a]	Alt.[b]	D-I[c]	Tot.[d]	I-D	Alt.	D-I	Tot.	I-D	Alt.	D-I	Tot.
1.	—[e]	—	—	—	—	—	—	—	4.4	—	4.9	9.4
2.	3.5	—	—	—	—	—	—	—	—	—	16.8	19.1
3.	12.8	—	6.2	21.8	—	—	—	—	—	5.5	—	—
4.	24.7	—	3.4	30.3	5.4	—	—	—	—	19.4	12.6	32.1
5.	—	—	—	—	—	—	—	—	—	—	—	—
6.	12.9	—	9.9	25.1	—	—	—	—	—	—	—	—
7.	—	—	—	—	—	—	—	—	—	—	—	—
8.	4.0	15.1	—	19.2	—	—	—	—	9.7	—	9.1	18.9
9.	—	36.9	37.9	74.8	—	—	3.7	—	—	—	—	—
10.	3.6	6.8	—	10.4	—	—	—	—	6.7	—	—	9.5
11	—	—	5.5	—	—	—	—	—	—	—	—	—
12.	—	—	36.1	39.4	—	—	—	—	—	—	4.0	—
13.	7.0	18.6	—	27.2	6.7	—	4.9	14.5	—	4.2	—	—
14.	11.2	—	16.8	30.4	—	5.8	—	8.4	—	—	—	—
15.	—	5.0	—	—	—	—	—	—	—	9.2	—	11.3
16.	—	11.0	6.3	18.0	—	—	—	—	—	—	—	—
17.	—	—	5.5	—	—	—	—	—	—	—	3.9	—
18.	14.1	10.5	28.9	53.4	—	—	—	—	18.2	7.3	4.7	30.2
19.	18.1	—	4.9	23.0	—	—	—	—	—	—	—	—
20.	7.5	16.0	5.9	29.4	—	—	—	—	—	—	—	—
21.	9.4	22.1	—	31.5	—	—	—	—	—	—	3.7	—
22.	5.9	12.5	13.5	31.9	—	—	10.1	14.5	—	3.7	—	—
23.	29.7	15.8	6.5	52.0	—	—	—	—	—	—	—	—
24.	—	11.8	15.2	27.6	—	15.0	5.0	20.1	—	—	—	—
25.	5.0	8.6	—	14.1	—	—	—	—	—	10.9	—	13.8
26.	27.6	13.8	—	41.8	9.4	11.8	7.5	28.7	3.7	11.6	—	16.6
27.	—	—	9.6	9.7	—	—	—	—	—	—	—	—
28.	—	—	—	—	—	—	—	—	6.2	22.8	—	29.3
29.	30.5	20.7	36.5	87.7	37.9	3.8	—	43.1	—	—	—	—
30.	—	14.8	—	19.1	—	—	—	—	—	—	—	—
31.	25.8	—	26.6	52.8	—	—	—	—	—	—	—	—
32.	15.1	24.7	—	41.0	4.0	—	3.5	7.7	—	9.5	—	10.1
33.	—	—	—	—	—	—	3.3	—	—	—	—	—
34.	6.9	11.0	—	19.3	—	—	—	—	—	4.0	3.7	7.8
35.	—	16.9	4.0	20.9	—	—	—	—	—	4.2	—	—
36.	—	16.3	12.3	28.8	—	4.2	—	8.6	21.9	11.7	—	35.1

[a] Indirect-direct order of presentation.
[b] Alternate order of presentation.
[c] Direct-indirect order of presentation.
[d] Totals are obtained by adding the independent Chi-squares in the three orders and using three degrees of freedom.
[e] Blank spaces represent non-significant Chi-square values; all printed Chi-square values are significant at least at the 5% level of confidence.

identity was somewhat greater. As before, D-1 to I-1 identity was considerably less than D-1 to I-2 identity.

2. When Table 11 is compared to Table 10 there seems to be only a moderate connection between the item identities made by the non-problem and by the problem children.

3. The subject matter of the questions showing the most consistent identity from Tests D-1 to D-2, I-1, and I-2 was as follows:

No. 10 Finds school a hard place to get along in.
No. 12 Finds it hard to speak out in class recitations.
No. 24 Gets angry.
No. 26 Dislikes school.

Three of these four questions concern school activities. This would indicate that the item differentiations of Table 9, which also were heavily loaded with school items, may have been caused by the problem children's reacting quite strongly to these items, and admitting negative symptoms on them not only on the direct but also the indirect questions. Their reactivity in this connection may have possibly been the result of their being already *known* as problem children, and therefore having little to lose by admissions of unfavorable attitudes towards school activities.

TABLE 11

Chi-square values indicating a significant degree of item identity from the items of one personality test to the equivalent items on three other personality tests taken by problem children (Three orders of presentation combined)

Item	Obtained Chi-square value		
	Test D-1 to D-2	Test D-1 to I-1	Test D-1 to I-2
1.	—*	—	—
2.	13.8	—	—
3.	—	—	—
4.	7.8	—	19.4
5.	9.4	—	10.2
6.	—	—	—
7.	—	—	—
8.	—	3.3	4.1
9.	8.4	—	—
10.	13.9	8.2	17.0
11.	—	—	—
12.	—	8.2	24.7
13.	18.7	—	—
14.	—	—	—
15.	—	—	4.4
16.	—	—	—
17.	—	—	—
18.	4.8	—	3.7
19.	13.8	—	—
20.	—	—	—
21.	—	—	—
22.	10.2	—	—
23.	9.2	—	5.9
24.	3.3	5.5	—
25.	19.4	—	9.7
26.	5.4	10.7	14.6
27.	—	—	—
28.	—	—	9.0
29.	24.4	—	—
30.	—	—	—
31.	11.9	—	9.9
32.	—	—	6.3
33.	—	—	3.7
34.	6.1	—	5.5
35.	—	—	—
36.	4.1	—	—

* Dashes represent non-significant Chi-square values; all printed Chi-square values are significant at least at the 5% level of confidence.

Problem 7. Did the degree of embarrassment which the subjects felt in answering the test items affect their responses?

Since analysis has shown that certain items on the test forms discriminated significantly between non-problem and problem children, and that item identity occurred on some of the questions, it may be asked whether the discriminations or item identities are related to the degree of embarrassment which the subjects experienced with the various questions. To answer this problem, the direct questions employed in the experiment (that is, Test D-1) were re-submitted, three months after the original administration, to one hundred of the original non-problem subjects. This time, however, the questions were given anonymously, and the children were requested to note whether they considered each one to be an embarrassing or a non-embarrassing one to which to respond. The results of this supplementary survey of the subjects' attitudes towards the direct questions are shown in Table 12.

From Table 12 it may be noted that, in general, the children found those questions embarrassing which asked whether other people were critical of them, whether they found it hard to get

TABLE 12

Percentage of one hundred non-problem school children finding personality test items embarrassing

Item	Subject Matter of Item	Percentage of children finding item embarrassing
1.	Would rather play by self than with other children	3*
2.	Finds it hard to make friends with other children	14
3.	Cries	25**
4.	Feels like running away from home	29**
5.	Finds it hard to do school work right	19**
6.	Is afraid of the dark	10
7.	Feels ill or weak	11
8.	Is afraid of being followed when walking on the street	11
9.	Feels other members of family are treated better than he	13
10.	Finds school a hard place to get along in	11
11.	Other children refuse to play with him	14
12.	Finds it hard to speak out in class recitations	8
13.	Feels that his teachers treat him badly or unfairly	13
14.	His feelings are easily hurt	26**
15.	Feels that nobody really loves him	12
16.	Feels tired during the daytime	4*
17.	Finds home a hard place to get along in	15**
18.	Feels tired when he wakes up in the morning	4*
19.	Is afraid during a thunder-storm	7
20.	Has headaches	1*
21.	Is bothered with pains in some part of his body	4*
22.	Is afraid of high places	1*
23.	Day-dreams	12
24.	Gets angry	8
25.	Has the habit of twitching his neck or face	5*
26.	Dislikes school	9
27.	People find fault with him	20**
28.	Parents blame or find fault with him	19**
29.	Bites his nails	17**
30.	Is afraid of fire	5*
31.	Has horrible dreams while sleeping	4*
32.	Finds it hard to become the leader in games with other children	22**
33.	Becomes gloomy or sad	8
34.	Becomes excited	10
35.	Stutters or stammers when talking	11
36.	Is afraid of his father	13

* The nine questions which were considered embarrassing by the fewest children.
** The nine questions which were considered embarrassing by the most children.

along with others, and whether they were sensitive to criticism. They found relatively non-embarrassing the questions which asked if they had physical ailments or were afraid of things like fire or high places. There are not many surprises in the tabulated results.

When the facts of Table 12 are viewed against those of Tables 9 and 10, no clear-cut results can be discerned. Thus, of the items in Table 9 which significantly differentiated non-problem from problem children, three of those of Test I-1 fell in the most-embarrassing, three in the least-embarrassing categories, and the other six in-between. Of the four I-2

discriminating items, one was least-embarrassing, and three in-between. Of the seventeen D-2 discriminating items, one was in the least-embarrassing area, five in the most embarrassing, and eleven in-between. Of the ten discriminating D-1 items, two fell in the least-embarrassing category, three in the most-embarrassing and five in-between. There is consequently little to indicate that the degree of embarrassment which the children said they felt affected the discriminative power of the test items.

Comparing the Table 12 values to those of the significant item identities listed in Table 10, results are similarly inconclusive. Thus, of the twenty-seven item identities from Test D-1 to D-2, eight fell in the most-embarrassing category, six in the least embarrassing, and thirteen in-between. Of the eight item identities from Test D-1 to I-1, one was most-embarrassing, three least-embarrassing and four in-between. Of the thirteen identities from Test D-1 to I-2, three fell in the most-embarrassing, three in the least-embarrassing, and seven in the in-between categories. It is evident, therefore, that according to the children's own attitudes towards degree of embarrassment experienced in answering the direct test items, there was no noticeable degree of correspondence between embarrassment or non-embarrassment on the one hand and item identity on the other.

Problem 8. How were the responses of the normal children affected by the subject matter of the items of the four tests?

To discover whether the subject matter of the items of the four tests affected the children's responses, the thirty-six questions of each form were divided into six categories, according to their content. These categories were: (1) questions relating to *home adjustment;* (2) to *social adjustment;* (3) to *school adjustment;* (4) to *fears;* (5) to *sensitivity and excitability;* and (6) to *psychosomatic symptoms.* Listed in Table 13 are the six categories, with the percentages of non-problem children responding negatively or unfavorably to each item. The problem subjects are not included since the small number of cases make their percentage figures meaningless.

Table 13 shows distinct differences among the average percentages of subjects responding negatively to the questions in each of the six subject categories. When tested with analysis of variance, these differences are significant for Tests I-1, I-2, and D-1, but not for D-2. None of the order differences are significant.

If the percentages in Table 13 are averaged by orders (since the order differences are not significantly different) the summary information shown in Table 13a is obtained. From Tables 13 and 13a it may be seen that:

1. On all four tests the subjects were least willing to respond negatively to the home adjustment and social adjustment questions and most willing to respond unfavorably to the psychosomatic and sensitivity and excitability questions.

2. There was some degree of correspondence between the manner in which the children responded to the six categories of questions on all four tests, but there was also some degree of discrepancy.

Additional information may be obtained by noting the nine most-checked and the nine least-checked items for each of the four tests. This information is shown in Table 14. From Table 14 it may be seen that on several of the questions—especially those concerned with the children's running away from home,

TABLE 13

Percentages of 221 non-problem children giving unfavorable responses to the items of four personality tests taken in three orders of presentation

Subject Matter of Item	Percentage of Children Giving Unfavorable Response											
	Test I-1			Test I-2			Test D-1			Test D-2		
	I-D[a]	Alt[b]	D-I[c]	I-D	Alt.	D-I	I-D	Alt.	D-I	I-D	Alt.	D-I
Home Adjustment:												
Feels like running from home	17	17	12	16	9	4	7	3	3	6	3	7
Feels family treated better	34	43	39	19	13	15	1	7	3	1	7	4
Finds home a hard place	17	16	21	3	7	7	3	4	1	1	5	4
Feels parents blame him	30	20	29	16	21	18	17	13	8	17	25	20
Is afraid of father	39	30	30	21	34	28	21	14	8	16	12	14
AVERAGE	29	25	26	15	17	14	10	8	5	8	10	10
Social Adjustment:												
Plays by himself	23	39	23	4	7	4	6	8	8	7	14	9
Finds it hard to make friends	26	31	28	7	19	11	3	8	5	3	9	5
Other boys won't play	23	39	27	17	12	9	0	4	4	0	5	1
Feels nobody loves him	14	13	14	8	10	12	11	14	1	6	3	8
Day-dreams often	33	35	32	41	35	36	13	14	18	13	14	5
Hard to become a leader	58	44	52	45	41	32	21	21	35	19	26	13
AVERAGE	30	34	30	21	21	18	8	12	12	8	12	10
School Adjustment:												
Hard to do school work	54	34	50	36	36	28	4	10	8	6	7	3
Finds school a hard place	33	25	18	36	36	32	11	9	7	3	5	5
Hard to speak in class	52	46	53	29	23	36	11	3	30	16	26	31
Teachers treat him badly	33	36	36	41	50	28	13	12	12	7	9	5
Dislikes school	41	34	35	60	42	48	21	14	12	11	7	11
AVERAGE	43	35	39	41	37	34	12	10	14	9	11	11
Fears:												
Is afraid of the dark	33	26	25	39	38	31	6	10	8	8	7	5
Afraid of being followed	26	27	19	30	27	32	7	9	8	19	9	11
Afraid during thunder-storm	36	40	26	30	22	26	6	12	8	4	7	11
Afraid of high places	53	46	60	50	58	46	7	17	18	13	19	16
Afraid of fire	80	61	54	40	39	35	6	12	3	10	18	11
Has horrible dreams	30	35	35	36	29	27	11	5	3	8	5	4
AVERAGE	43	39	37	38	36	38	7	11	8	10	11	10
Sensitivity & Excitability:												
Cries easily	37	30	48	56	25	46	10	12	9	11	12	10
Feelings easily hurt	36	39	50	41	36	35	10	15	25	8	10	15
Gets angry	36	19	38	80	74	59	24	23	30	23	26	26
People find fault with him	41	30	23	21	17	20	4	7	12	8	9	16
Feels gloomy or sad	40	46	18	30	27	19	3	3	7	3	19	5
Becomes excited	47	39	44	73	54	42	29	27	27	11	5	18
AVERAGE	42	34	37	50	39	36	13	15	18	11	13	14
Psychosomatic Symptoms:												
Feels ill or weak	64	46	50	37	30	18	4	3	3	6	4	16
Tired during the daytime	54	50	46	23	27	18	6	8	7	4	8	4
Tired when waking	80	82	72	66	68	59	37	39	39	26	26	9
Has headaches	66	70	41	52	27	35	22	17	13	8	10	9
Pains in body	71	52	68	33	23	19	13	5	15	7	8	4
Twitches neck or face	27	34	36	7	18	9	1	17	5	4	8	8
Bites nails	54	52	50	78	60	64	33	29	38	23	15	32
Stutters or stammers	67	57	55	27	29	22	1	7	15	3	7	8
AVERAGE	60	55	52	40	35	30	14	16	17	13	11	11

a—Indirect-Direct Order b—Alternate Order c—Direct-Indirect Order.

finding it hard to make friends, tiredness when waking in the morning, and biting their nails—there was a definite tendency for the subjects to give the same kind of response to all four tests. But on a few of the questions—specifically, those concerned with the subjects' feeling their parents blamed them, being afraid of being followed, feeling gloomy or sad, and feeling ill or weak—they freely gave unfavorable responses to one form of the test but rarely gave them to another form.

Table 14 confirms the information of Tables 13 and 13a: to wit, that non-problem subjects, on all four tests, ad-

TABLE 13a

Average percentages of non-problem children giving unfavorable responses to six categories of questions on four personality tests (three orders of presentation combined)

Category of Questions	Average Percentage Giving Unfavorable Response			
	Test D-1	Test D-2	Test I-2	Test I-2
Home Adjustment	7.7	9.3	26.8	15.3
Fears	8.7	10.3	39.6	37.3
Social Adjustment	10.7	10.0	31.3	20.0
School Adjustment	12.0	10.3	39.0	37.3
Sensitivity and Excitability	15.3	12.7	37.7	41.7
Psychosomatic Symptoms	15.7	11.7	55.7	35.0

mitted with relative rarity that they were unadjusted to their home and that they were socially inadequate; but more freely admitted that they were sensitive and excitable and that they possessed psychosomatic symptoms. This pattern of response may be attributable to the fact that (1) the children really did possess fewer unfavorable home and social symptoms than sensitivity and psychosomatic ones; or (2) they felt less self-criticism in admitting the latter than the former kind of symptoms. To check this second possibility, the information in Table 12, showing the degree of embarrassment felt by non-problem children in answering the direct items, may be compared with the data of Table 14. It is then found that of the nine questions considered most-embarrassing by the children, three were among the least-checked items in one or more of the tests, four were among the most-checked, one was least-checked in two tests and most-checked in one, and one was neither among the most- or least-checked items. Of the nine items considered least-embarrassing by the children, five were among the most-checked items on one or more tests, and four were among the least-checked. Apparently, there was not a high degree of correspondence between the children's finding specific items embarrassing, on the one hand, and giving unfavorable answers to them on the other hand. The observed differences among the six categories of questions may therefore be attributable to actual differences in the subjects' possession of the various traits rather than the result of their responding more unfavorably to the less self-critical items. Certainly, it is easy to believe, as the data of Table 14 indicate, that non-problem children less often feel like running away from home and less often find it hard to make friends with other children than they feel tired when waking in the morning and than they bite their nails.

Problem 9. What was the effect of the children's answering certain direct and indirect questions out of context as against answering identical questions as part of the regular test battery?

The test booklet used in the investigation contained three sample questions on their covers. The sample D-1 question was: "I get angry VERY OFTEN PRETTY OFTEN SELDOM NEVER." The sample I-1 quesion was: "Children who often day-dream are VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER." The sample I-2 question was: "VERY MANY PRETTY MANY FEW NO children are afraid of high places." No sample D-2 question was employed, since its form was similar to that of the D-1

TABLE 14

The nine most frequently checked and nine least frequently checked items on four personality tests taken by 221 non-problem children (orders combined)

Subject Matter of Item	Most-Checked Test				Least-Checked Test			
	I-1	I-2	D-1	D-2	I-1	I-2	D-1	D-2
Home Adjustment:								
Feels like running away from home					x	x	x	x
Feels family treated better than he is						x	x	x
Finds home a hard place						x	x	x
Feels parents often blame him				x	x	x		
Is afraid of his father				x				
Social Adjustment:								
Plays by himself					x	x		
Finds it hard to make friends					x	x	x	x
Other boys refuse to play with him						x	x	x
Feels that nobody loves him					x	x		
Day dreams often			x					
Finds it hard to become a leader	x							
School Adjustment:								
Finds it hard to do school work								x
Finds school a hard place					x			x
Finds it hard to speak in class			x	x				
Feels teachers treat him badly		x						
Dislikes school		x	x					
Fears:								
Is afraid of the dark					x			
Is afraid of being followed when walking				x	x			
Is afraid during a thunder-storm					x			
Is afraid of high places	x	x		x				
Is afraid of fire	x							
Has horrible dreams							x	x
Sensitivity and Excitability:								
Cries easily		x						
Feelings are easily hurt			x					
Often gets angry		x	x	x				
Feels people find fault with him								
Feels gloomy or sad				x			x	
Becomes excited		x	x					
Psychosomatic Symptoms:								
Feels ill or weak	x						x	
Feels tired during the daytime							x	x
Is tired when waking in the morning	x	x	x	x				
Has headaches	x	x	x					
Has pains in his body	x							
Twitches his neck or face					x			
Bites his nails	x	x	x	x				
Stammers or stutters	x							

questions. The children were not specifically requested to answer these sample questions, but 101 of the non-problem children voluntarily did so, presumably as aids to themselves. Accordingly, since the sample questions were also contained in the regular context of the tests, the two responses of these 101 subjects to each of the three sample questions were compared to see if the contextual one was answered in (1) a more favorable, (2) a less favorable, or (3) an unchanged

manner from the non-contextual responses. The results of the comparisons are shown in Table 15.

From Table 15 the following points may be noted:

1. Almost one-fourth of the children answering the sample, non-contextual direct question changed their contextual response to a more favorable one. At the same time, only 4% changed their response less favorably. If it is assumed that the non-contextual answers were the more truthful ones—since they were presumably given when the children were caught off their guard, and given to questions on which they felt they would not be marked—there is some evidence that sophistication occurred as a result of the test situation. And since the particular direct questioned involved here—that asking whether the children often got angry—was one which, according to Tables 12 and 14, the subjects freely responded to unfavorably, and according to Table 13 did not deem particularly embarrassing, there is some reason to believe that the test sophistication shown here also would prevail for most other direct questions.

2. The direction of the changes made in the I-2 sample question were similar to those of the D-1 question, but not so pronounced as to evidence of test sophistication. Some 18% of the children changed their I-2 response in a more favorable direction; but 10% also changed it in a less favorable manner. There seems to be nothing conclusive about these changes.

3. The Test I-1 sample question produced results opposite to those of the other two questions. This time, the

TABLE 15

Changes made by 101 non-problem children in answering certain personality test questions first out of context and then as part of a regular test battery

Direction of Change in Response	Number and percentage of children making change							
	Indirect-Direct Order		Alternate Order		Direct-Indirect Order		Combined Orders	
	N	%	N	%	N	%	N	%
D-1 Question (gets angry)								
More favorable	4	13	12	29	8	25	24	24
Less favorable	1	3	1	2	2	6	4	4
Unchanged	22	84	29	69	22	69	73	72
Total	27	100	42	100	32	100	101	100
I-1 Question (day-dreams often)								
More favorable	7	23	6	14	4	12	17	17
Less favorable	4	13	13	31	9	28	26	26
Unchanged	16	24	23	55	19	60	58	57
Total	27	100	42	100	32	100	101	100
I-2 Question (is afraid of high places)								
More favorable	4	13	6	14	8	25	18	18
Less favorable	5	17	4	10	1	3	10	10
Unchanged	18	70	32	76	23	72	73	72
Total	27	100	42	100	32	100	101	100

children changed their answers to more favorable responses 17% of the time and to less favorable ones 26% of the time. All told, no less than 43% of the subjects changed their contextual response to this item. Several possible explanations for this unusual response pattern present themselves: (1) the children never quite figured out what were the favorable answers to this question, and were therefore as willing to change their answers one way as the other; (2) what the scoring system assumed to be "favorable" answers to this test did not prove to be so empirically; (3) the subject matter of the question—that asking about the children's day-dreams—was especially conducive to unstable responses. Since none of these explanations, individually considered, seems conclusive, it is possible that a combination of them accounted for the observed changes.

Because of the differences in subject matter of the three sample questions, it is not possible directly to compare the changes made on one to those made on the other two. However, since the three are fairly equivalent in the sense that none of them was found to be very embarrassing by the children (Table 12) and all tended to be among the most-checked responses (Table 14), there are some indications here that, given the scoring procedures used in this investigation, the D-1 question was somewhat transparent to the subjects, the I-2 question less transparent, and the I-1 question non-transparent. This agrees with the item identity findings, which showed that the I-2 questions were more closely related to the D-1 items than were the I-1 questions.

RESUME OF THE FINDINGS

The major findings of the experiment, classified according to the main variables involved in it, are briefly listed below.

Tests

1. The mean scores of the direct tests (D-1 and D-2) do not differ significantly.

2. The average scores of the indirect tests (I-1 and I-2) are significantly different.

3. The mean scores of the direct tests (D-1 and D-2) differ significantly from those of both indirect tests (I-1 and I-2).

4. The observed mean score differences are not attributable to lack of internal consistency or lack of item stability of the tests.

5. The intercorrelations between Tests D-1 and D-2 are reasonably high; between I-1 and I-2, fairly low; and between the direct and indirect tests, not too high.

6. Many item identities exist between Test D-1 and D-2; a good many between D-1 and I-2; but relatively few between D-1 and I-1.

7. The subjects give significantly more unfavorable responses to certain kinds of questions—*e.g.*, psychosomatic ones—than they do to other kinds—*e.g.*, home adjustment or social adjustment items.

8. Subjects taking specified test items as part of the regular test battery respond differenly than when taking identical items out of test context.

9. The degree of embarrassment the subjects say they experience in answering the direct questions does not affect the power of the items to discriminate between the two groups or the item identities of the questions.

Groups

10. All the obtained group differences are to be cautiously viewed, since the problem subjects are few in number and

are probably non-homogeneously subgrouped.

11. Significant differences between problem and non-problem subjects' average scores are evident only for Test D-2, and not for Tests D-1, I-1, or I-2.

12. Few items discriminate significantly between problem and non-problem groups on any of the four tests, though the direct tests contain more item discriminations than do the indirect. Different specific items prove discriminating on the direct and the indirect tests.

13. The items which discriminate significantly between problem and non-problem groups are largely concerned with school activities.

Orders of Presentation

14. The order of presentation of the direct and the indirect test questions has a significant effect upon:

a. Average test scores.
b. Test score intercorrelations.
c. Item discriminations between problem and non-problem groups.
d. Biserial correlations between total test and item scores.
e. Percentages of subjects responding unfavorably to different kinds of questions.

Interactions

15. None of the first-order or second-order interactions among tests, groups, or orders of presentation are statistically significant.

CHAPTER IV

DISCUSSION

SINCE statistically significant differences were found for tests, orders of presentation, and groups used in the experiment, it may now be asked why they occurred and what they mean.

INTERPRETATION OF THE TEST DIFFERENCES

The indirect tests employed in the experiment differed significantly from each other as well as from the direct tests; and these differences were maintained for both problem and non-problem subjects. They may be attributable to differences between the indirect and direct tests because of (1) format or (2) substance.

It seems unlikely that the test differences were the result of format differences because: (1) deliberate change in phrasing, but avoidance of change in meaning, did *not* bring about significant differences between Tests D-1 and D-2. (2) The formats of Tests I-1 and I-2 differed from each other; yet, *both* I-1 and I-2 mean scores were significantly different from D-1 and D-2 scores. (3) Changing the order of presentation of the tests, which somewhat changed their appearance or format, had no significant effect on the distinct differences found between the direct and indirect test forms.

Since there is little reason to believe that the format of the tests was responsible for the observed mean score differences, these would seem to be attributable to differences in test substance.

INTERPRETATION OF THE GROUP DIFFERENCES

It was found, by statistical analysis, that the problem children reacted differently than the non-problem subjects to Test D-2; but that there were no significant differences between the mean scores of the two groups on Tests D-1, I-1, and I-2. The most plausible reasons for these diverse group differences seem to be as follows:

1. Tests D-2 and I-1, which most successfully differentiated the two groups (the latter displaying differences which were almost significant), seem to have been the tests with which the subjects had the greatest difficult determining the "good" or "correct" answers. Some of the D-2 questions were awkwardly worded and hence liable to misinterpretation; while few of the I-1 questions—as the item identity data revealed—could be readily identified by the subjects with their D-1 analogues. Consequently, the problem children, who were found to have lower intelligence quotients than the non-problem subjects, may have answered the D-2 and the I-1 items less sophisticatedly than did the latter group: thereby producing the observed differences.

2. The group differences could have been caused by lack of homogeneity among the subgroups of problem children used in the experiment. These problem subjects, not being the most acute and segregable kind of cases, necessarily consisted of a heterogeneous population which probably included several clinical subgroups. From such a heterogeneous grouping, a sample as small as that employed in the experiment could easily have varied significantly from one order of presentation to another, thus causing some inconsistencies

in the findings. Observation of Table 1 (p. 7) tends to bear out this possibility by showing that while the order of presentation scores for the non-problem children ran in a consistent direction for all tests, those for the problem children were either inconsistent or disproportional from order to order. That is why sophistication effects found for the non-problem group were not also found for the problem subjects. The internal consistency coefficients (Table 5, p. 12) also show evidence of inconsistency in the problem children's responses, and this inconsistency again could have been caused by differences within the sub-groups of problem subjects. Since, throughout the entire experiment, the non-problem children, whose sub-groups always included approximately seventy cases, showed fairly consistent findings, and the problem children, whose sub-groups included approximately thirteen cases, often showed inconsistent and contradictory findings, there must be a strong suspicion that the obtained differences between the mean scores of the two groups resulted from unusual heterogeneity in the makeup of the problem subgroups. Consequently, not too much emphasis may be placed on the observed group differences until they are verified by further experimentation. The obtained test and order of presentation differences, on the other hand, stand up much better under statistical analysis.

INTERPRETATION OF THE ORDER OF PRESENTATION DIFFERENCES

The obtained mean score differences on the three orders of presentation used in the experiment seem to be clearly explicable in terms of test sophistication. Bolstering the argument that one presentation order enabled the subjects to become more sophisticated on the tests than did the other orders are the following points:

1. The hypothesis of test sophistication would lead to the expectation that different orders of presentation would produce greater differences on the indirect than on the direct tests, since the latters' questions were so openly stated as to reveal their intent no matter what presentation order was employed. This was precisely what occurred, with order differences being significant only on Tests I-1 and I-2, and not on D-1 or D-2.

2. If sophistication effects existed, they should have been revealed by the subjects' receiving higher (or less favorable) scores on the indirect tests when they answered all the indirect questions before the directly stated ones, than when they first answered the direct and then the indirect questions. This is exactly what the experimental data reveals them to have done.

3. If the sophistication hypothesis were tenable, it would be expected that subjects taking the alternate order of presentation would tend to receive mean test scores in-between those taking the indirect-direct and direct-indirect orders. This, again, was revealed by the data.

4. The sophistication hypothesis would leave the probability of some degree of sophistication taking place on the direct tests, since subjects answering these tests last (that is, taking the indirect-direct order) might be made somewhat test-wise. The obtained scores show that some (though a statistically non-significant) degree of sophistication did occurr on the direct tests, and that it was precisely in the expected direction.

On all counts, therefore, the hypothesis seems to be upheld; and it would definitely appear that the observed differ-

ences produced by the orders of presentation were attributable to test sophistication. This finding, and the obvious implications which go with it, may be said to be the most important discovery of the investigation.

Chapter V

Summary and Conclusions

The purpose of the present experiment was to study the effects of using direct and indirect phrasing in personality questionnaires. The specific hypothesis to be tested was that if conventional personality test schedules employing direct questions, as well as indirect test forms derived from the direct items, are given in a single battery to a group of behavior problem and a group of non-problem school children: (a) the indirect test scores will be significantly different from the direct scores; (b) groups of behavior problem and non-problem children will react differently to the indirect and direct tests; and (c) presentation of the questions in three different orders—an indirect-direct, an alternate, and a direct-indirect order—will have a distinct effect on the degree of test sophistication shown by the children in their responses.

Summary of the Experiment

In order to test the experimental hypothesis, these materials, subjects, and experimental design were employed:

Materials. A basic personality test form of 36 items was constructed by taking the most frequently used questions on eight standard personality questionnaires for children. Two direct forms of this test were made: (1) with all 36 questions asking about a negative symptom (*e.g.*, "I cry VERY OFTEN PRETTY OFTEN SELDOM NEVER."); (2) with all 36 questions asking about a positive symptom (*e.g.*, "I am free from crying VERY OFTEN PRETTY OFTEN SELDOM NEVER."). Two indirect forms of the same 36 questions were constructed: (1) with all questions asking how queer the respondents considered other children who possessed negative symptoms (*e.g.*, "Children who often cry are VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER."); (2) with all questions asking the respondents how many other children there were who possessed negative symptoms (*e.g.*, "VERY MANY PRETTY MANY FEW NO children often cry."). In addition, the Pintner Intermediate Test of Mental Ability was administered to all subjects.

Subjects: The experimental groups consisted of forty behavior problem boys and 221 non-problem boys from regular seventh and eighth grade classes of three New York City junior high schools.

Experimental Design. The four test forms were bound together in sixteen-page booklets, each page containing nine questions from one of the tests. The booklets were bound so that the tests could be administered, in a single setting, in three different orders of presentation: (1) an indirect-direct order—all eight pages of the indirect questions followed by the eight pages of the direct questions; (2) an alternate order—a page of direct questions alternated with a page of indirect ones; and (3) a direct-indirect order—all eight pages of the direct questions followed by the eight pages of the indirect ones. The booklets were distributed so that one-third of the problem and one-third of the non-problem subjects received each order of presentation. Problem and non-problem children were tested in their regular home-room classes, with no segregation of the groups taking place.

Conclusions and Implications

The results of the experiment substantially support its hypothesis, the

main findings being as follows:

1. The mean scores of the direct tests do not differ significantly from each other; but those of the indirect tests do. The mean scores of the two direct tests differ significantly from those of both indirect tests.

2. Intercorrelations between the two direct tests are reasonably high; and many item identities exist between these two tests. Intercorrelations between the two indirect tests are fairly low. Intercorrelations between the direct and indirect tests are not too high; and item identities from the direct to the indirect tests are not too many. One indirect test (I-2) is much more closely related to the direct tests than is the other indirect test (I-1).

3. The subjects give significantly more unfavorable responses to certain kinds of questions, especially psychosomatic ones, than they do to other kinds, especially home and social adjustment items.

4. The subjects react quite differently to questions taken out of the test context than they do to those taken as part of the regular test battery.

5. Significant differences between problem and non-problem subjects' mean scores are evident only for one of the direct tests (D-2) and not for the other direct test or for either of the direct tests. Observed differences between problem and non-problem groups of subjects are probably atributable to the small number and non-homogeneous subgrouping of the problem children and are therefore to be cautiously viewed and interpreted.

6. Few items discriminate significantly between the problem and non-problem groups on any of the four tests, though the direct tests contain more item discriminations than do the indirect. Different specific items prove discriminating on the direct and indirect tests. The items which discriminate best between problem and non-problem groups are largely concerned with school activities.

7. The order of presentation of the direct and the indirect test questions has a significant effect upon average test scores, test score intercorrelations, item discriminations between problem and non-problem groups, biserial correlations between total test and item scores, and percentages of subjects responding unfavorably to different kinds of questions.

The main implications to be drawn from the experiment are presented below. In making them, it must be emphasized that the present study, being based on the conventional types of questionnaires which are now widely used, is doubtlessly limited by some of the defects which are inherent in such tests. A particular limitation of the study is its use, for its problem subjects, of a smaller group of children than would have been statistically desirable. Consequently, the following conclusions and implications must be taken as tentative, and as requiring for their final confirmation considerable further experimentation.

1. The order of presentation of indirect and direct personality questions administered in a single test battery may decisively effect test results. If test-makers are to construct instruments that will encourage the least sophisticated, most truthful responses by their subjects, they must make certain that all indirect items are printed before the direct ones.

2. Phrasing is an important consideration in the construction of personality test items. Each mode of indirectly phrasing a schedule of questions may result in the creation of an instrument

significantly different from that of other indirectly phrased questionnaires.

3. Indirect personality tests seem to differentiate problem from non-problem groups of children no better than do direct test forms. Certain specific questions, however, are more discriminative when indirectly than directly phrased, so that optimum results may possibly be obtained by combining both kinds of questions in a single test.

4. The kinds of questions which appear best to distinguish problem from non-problem groups of children are those specifically concerned with school activities.

5. Much more experimental work must be done with personality questionnaires before they can be expected to be satisfactory instruments for the diagnosis of behavior problem difficulties.

BIBLIOGRAPHY

1. ADAMS, C. R. and LEPLEY, W. M. *The Personal Audit.* State College, Pennsylvania.
2. BARTLETT, M. S. Some examples of statistical methods of research in agriculture and applied biology. *Suppl. to J. Roy. Stat.*, 1937, *4*, 137-83.
3. BERNREUTER, R. G. *The Personality Inventory.* Stanford: Stanford University Press, 1935.
4. ELLIS, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, *43*, 385-440.
5. ELLIS, A. Personality questionnaires. *Review of Educ. Res.*, 1947, *17*, 53-63.
6. FISHER, R. A. *Statistical methods for research workers.* Edinburgh: Oliver and Boyd, 1925-1944.
7. HATHAWAY, S. R. and MCKINLEY, J. C. A Multiphasic Personality Schedule (Minnesota): I. Construction of the schedule. *J. Psychol.*, 1940, *10*, 249-54.
8. HILDRETH, GERTRUDE. *Personality and Interest Inventory, Elementary Form.* Bureau of Publications, Teachers College, Columbia University, New York, 1936.
9. HUMM, D. G. and WADSWORTH, G. The Humm-Wadsworth Temperament Scale. *J. appl. Psychol.*, 1935, *92*, 163-200.
10. JACKSON, R. W. B. and FERGUSON, G. A. *Studies on the reliability of tests.* Department of Educational Research, University of Toronto, 1941.
11. KUDER, G. F. and RICHARDSON, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, *2*, 151-60.
12. LAIRD, D. A. *Personal Inventory.* Hamilton, New York: Hamilton Republican, 1925.
13. LOOFBOUROW, G. C. and KEYS, N. *Personal Index.* Educational Test Bureau, Minneapolis, 1933.
14. MALLER, J. B. *Character Sketches.* Bureau of Publications, Teachers College, Columbia University, New York, 1932.
15. MCKINLEY, J. C. and HATHAWAY, S. R. The Minnesota Multiphasic Personality Inventory. V. Hysteria, hypomania and psychopathic deviate. *J. appl. Psychol.*, 1944, *28*, 153-74.
16. PINTNER, R. and OTHERS. *Pintner Intermediate Test. Form A.* World Book Company, Yonkers-on-Hudson, New York, 1928.
17. PINTNER, R. and OTHERS. *Aspects of Personality.* World Book Company, Yonkers-on-Hudson, New York, 1937.
18. ROGERS, C. R. *A Test of personality adjustment.* Association Press, New York, 1931.
19. SHIPLEY, W. C. and OTHERS. The Personal Inventory—its derivation and validation. *J. Clin. Psychol.*, 1946, *4*, 318-22.
20. SNEDECOR, G. W. *Calculation and interpretation of analysis of variance.* Collegiate Press, Ames, Iowa, 1934.
21. SNEDECOR, G. W. *Statistical methods.* (Fourth ed.) Collegiate Press, Ames, Iowa, 1946.
22. SPENCER, D. *Fulcra of conflict; a new approach to personality measurement.* World Book Co., Yonkers-on-Hudson, New York, 1938.
23. SWEET, L. *The measurement of personality attitudes in younger boys.* Association Press, New York, 1929.
24. SYMONDS, P. M. and BLOCK, VIRGINIA L. *Student questionnaire.* Bureau of Publications, Teachers College, Columbia University, New York, 1932.
25. TIEGS, E. W. and OTHERS. *California test of personality—Intermediate Form A.* California Test Bureau, Los Angeles, 1942.
26. TUCKMAN, J. Unpublished personality test. On file with Prof. Irving Lorge, Teachers College, Columbia University, New York.
27. WASHBURNE, J. N. *Washburne social-adjustment inventory.* World Book Co., Yonker-on-Hudson, New York, 1940.
28. WOODWORTH, R. S. and MATHEWS, ELLEN. *Personal data sheet.* Stoelting, Chicago, 1924.
29. YATES, F. Contingency tables involving small numbers and the Chi-square test. *Suppl. to the J. Royal Stat. Soc.*, 1934, 1, 217-35.

APPENDIX

THE TESTS AND INSTRUCTIONS EMPLOYED IN THE EXPERIMENT

Instructions

(To be read aloud to the Class)

Do not open your booklet until you are given the signal to do so. We would like your help in finding out what school children think about themselves and what they think about other children. Please answer the questions as honestly as you can, so that we may find out what school children, including yourself, are really like. Of course, this is not a test. These questions have nothing to do with your school work. Your teacher will not see them. You will not be marked on them.

Look at the top of the page in front of you. Fill it out now. Write in your name. Write today's date. Write down how old you were at your last birthday. Circle whether you are a boy or a girl. Put down your official class. Write the number of this school.

Do not open this booklet until you are given the signal to go. When you do open it, you will find that it contains three different kinds of questions. Samples of each one of these three kinds of questions are printed on the page before you. You may now look at these sample questions. First look at Question 1. Read it to yourself as I read it to you. It says: I get angry VERY OFTEN PRETTY OFTEN SELDOM NEVER." This question asks something about *yourself*. Answer it truthfully about yourself. Some of you may get angry VERY OFTEN, some of you PRETTY OFTEN, some SELDOM, and some of you NEVER. One of these descriptions certainly fits you. So draw a circle around the one which does describe how angry you get. Remember: only draw a circle around ONE answer for each question.

Now look at Question 2. This question reads: "VERY MANY PRETTY MANY FEW NO children are afraid of high places." It asks about *other* children whom you know. We want your opinion about how many of these other children are afraid of high places. Again, circle only *one* of the four answers to this kind of question.

Now look at Question 3, which reads: "Children who often day-dream are VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER." This question, again, asks about other children. We want your opinion about children who often day-dream, and want you to let us know if you personally think they are VERY QUEER, A LITTLE QUEER, or NOT AT ALL QUEER. This is not a test, remember, and you will not be marked on it.

You can see, then, that there are three different kinds of things you are to do in this booklet. One kind of question asks about yourself, and the other two kinds ask your opinions about other children. In all three kinds of questions, though, you are to draw a circle around only ONE of the four possible answers. Carefully read each question and answer it as honestly as you can. Do not skip any of them. The booklet has sixteen pages in it, but it should not take too long to answer all its questions. There is no speed limit, but work as fast as you can.

Your truthful answers on this booklet will be a great help to us in finding out what school children are really like, so be sure to answer every single question carefully. This has nothing to do with your school work, and your teacher will not even look at your papers.

Now open the booklet and make sure that there are exactly sixteen white pages of questions. If there are, go right ahead and answer all the questions. If there are not, ask your teacher for a new booklet.

Test D-1 Questions

I would rather play by myself than play with other children VERY OFTEN PRETTY OFTEN SELDOM NEVER

I find it hard to make friends with other children VERY OFTEN PRETTY OFTEN SELDOM NEVER

I cry	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel like running away from home	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find it hard to do my school work right	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am afraid of the dark	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel ill or weak	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am afraid of being followed when I am walking on the street	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
Other members of my family are treated better than I am treated	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find school a hard place to get along in	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
Other children refuse to play with me	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find it hard to speak out in class recitations	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel that my teachers treat me badly or unfairly	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
My feelings are easily hurt	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel that nobody really loves me	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel tired during the daytime	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find home a hard place to get along in	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel tired when I wake up in the morning	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am afraid during a thunderstorm	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I have headaches	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am bothered with pains in some part of my body	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am afraid of high places	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I day-dream	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I get angry	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I have the habit of twitching my neck or face	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I dislike school	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
People find fault with me	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
My parents blame me or find fault with me	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I bite my nails	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am afraid of fire	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I have horrible dreams while sleeping	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find it hard to become a leader in games with other children	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I become gloomy or sad	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I become excited	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I stutter or stammer when talking	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am afraid of my father	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER

Test D-2 Questions

I find it easy to speak out in class recitations	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am unafraid of my father	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel well and strong	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am unafraid of fire	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am happy	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I remain calm and unexcited	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER

UNIVERSITY OF MICHIGAN LIBRARIES

Other children are willing to play with me	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find school an easy place to get along in	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am treated as well as other members of my family are treated	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I like school	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from crying	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I keep from biting my nails	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
People praise me	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from anger	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am unafraid of the dark	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from feelings of wanting to run away from home	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from having my feelings easily hurt	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am unafraid during a thunderstorm	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find home an easy place to get along in	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel very rested when I wake up in the morning	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from day-dreaming	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from pains in some part of my body	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find it easy to become a leader in games with other children	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am unafraid of high places	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from stuttering or stammering when talking	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
My parents praise me	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
While sleeping I am free from horrible dreams	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from the habit of twitching my neck or face	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find it easy to make friends with other children	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I am free from having headaches	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel full of energy during the day	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel that my teachers treat me well and fairly	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I find it easy to do my school work right	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I have no fear of being followed when I am walking on the street	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I feel that people really love me	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER
I would rather play with other children than play by myself	VERY OFTEN	PRETTY OFTEN	SELDOM	NEVER

Test I-1 Questions

Children who feel tired when they wake in the morning are	VERY QUEER	PRETTY QUEER	A LITTLE QUEER	NOT AT ALL QUEER
Children who are afraid of fire are	VERY QUEER	PRETTY QUEER	A LITTLE QUEER	NOT AT ALL QUEER
Children who find it hard to make friends with other children are	VERY QUEER	PRETTY QUEER	A LITTLE QUEER	NOT AT ALL QUEER
Children who often feel ill or weak are	VERY QUEER	PRETTY QUEER	A LITTLE QUEER	NOT AT ALL QUEER

Children who feel that other members of their family are treated better than they are treated are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who dislike school are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children whom other children refuse to play with are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who find it hard to speak out in class recitations are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who would rather play by themselves than play with other children are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who often have headaches are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who often become gloomy or sad are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who are afraid of their father are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who feel that people often find fault with them are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who are bothered with pains in some part of their body are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who feel that nobody really loves them are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who are afraid of being followed when walking on the street are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who often feel like running away from home are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who find it hard to become a leader in games with other children are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who often cry are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who are afraid of high places are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who find it hard to do their school work right are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who find school a hard place to get along in are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who often get angry are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who feel that their parents blame them or find much fault with them are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who often become excited are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who often have horrible dreams while sleeping are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who find home a hard place to get along in are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who often day-dream are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who are afraid of the dark are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who are afraid during a thunder-storm are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who bite their nails are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

UNIVERSITY OF MICHIGAN LIBRARIES

Children who often feel tired during the daytime are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children whose feelings are easily hurt are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who feel that their teachers treat them badly or unfairly are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who stutter or stammer when talking are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Children who have the habit of twitching their neck or face are
VERY QUEER PRETTY QUEER A LITTLE QUEER NOT AT ALL QUEER

Test I-2 Questions

VERY MANY	PRETTY MANY	FEW	NO	children feel tired when they wake up in the morning.
VERY MANY	PRETTY MANY	FEW	NO	children bite their nails.
VERY MANY	PRETTY MANY	FEW	NO	children often get angry.
VERY MANY	PRETTY MANY	FEW	NO	children are afraid of fire.
VERY MANY	PRETTY MANY	FEW	NO	children find it hard to speak out in class recitations.
VERY MANY	PRETTY MANY	FEW	NO	children often feel tired during the daytime.
VERY MANY	PRETTY MANY	FEW	NO	children often become gloomy or sad.
VERY MANY	PRETTY MANY	FEW	NO	children are afraid of high places.
VERY MANY	PRETTY MANY	FEW	NO	children are afraid of being followed when they are walking on the street.
VERY MANY	PRETTY MANY	FEW	NO	children feel that their teachers treat them badly or unfairly.
VERY MANY	PRETTY MANY	FEW	NO	children are bothered with pains in some part of their body.
VERY MANY	PRETTY MANY	FEW	NO	children would rather play by themselves than play with other children.
VERY MANY	PRETTY MANY	FEW	NO	children find home a hard place to get along in.
VERY MANY	PRETTY MANY	FEW	NO	children often day-dream.
VERY MANY	PRETTY MANY	FEW	NO	children are afraid of their fathers.
VERY MANY	PRETTY MANY	FEW	NO	children find it hard to make friends with other children.
VERY MANY	PRETTY MANY	FEW	NO	children have other children refuse to play with them.
VERY MANY	PRETTY MANY	FEW	NO	children find it hard to do their school work right.
VERY MANY	PRETTY MANY	FEW	NO	children have the habit of twitching their neck or face.
VERY MANY	PRETTY MANY	FEW	NO	children feel that nobody really loves them.
VERY MANY	PRETTY MANY	FEW	NO	children often feel like running away from home.
VERY MANY	PRETTY MANY	FEW	NO	children often feel ill or weak.
VERY MANY	PRETTY MANY	FEW	NO	children often cry.
VERY MANY	PRETTY MANY	FEW	NO	children often have horrible dreams while sleeping.
VERY MANY	PRETTY MANY	FEW	NO	children find school a hard place to get along in.
VERY MANY	PRETTY MANY	FEW	NO	children feel that other members of their family are treated better than they are treated.

VERY MANY	PRETTY MANY	FEW	NO	children find it hard to become a leader in games with other children.
VERY MANY	PRETTY MANY	FEW	NO	children feel that their parents blame them or find much fault with them.
VERY MANY	PRETTY MANY	FEW	NO	children stutter or stammer when talking.
VERY MANY	PRETTY MANY	FEW	NO	children feel that people find much fault with them.
VERY MANY	PRETTY MANY	FEW	NO	children often become excited.
VERY MANY	PRETTY MANY	FEW	NO	children are afraid of the dark.
VERY MANY	PRETTY MANY	FEW	NO	children dislike school.
VERY MANY	PRETTY MANY	FEW	NO	children are afraid during a thunder-storm.
VERY MANY	PRETTY MANY	FEW	NO	children often have headaches.
VERY MANY	PRETTY MANY	FEW	NO	children's feelings are easily hurt.

UNIVERSITY OF MICHIGAN LIBRARIES